

Содержание

От издательства	14
Отзывы о книге	15
О переводчике	18
Вступительное слово от создателя Polars Ричи Винка	19
Введение	22
Об изображении на обложке	27
 Глава 1. Введение в Polars	 29
Что из себя представляет Polars	30
Ключевые особенности	30
Основные концепции	30
Преимущества	31
Почему вам стоит использовать Polars	32
Быстродействие	32
Удобство использования	33
Популярность	33
Рациональное использование ресурсов	34
Polars в сравнении с другими пакетами для обработки данных	35
Почему мы остановились на реализации Polars для Python	36
Организация книги	37
Пример на основе процесса ETL	38
Извлечение	39
Бонус: визуализация районов Нью-Йорка и расположения станций велосипедов	45
Преобразование	48
Бонус: визуализация поездок по дням и округам	53
Загрузка	55
Бонус: ускорение за счет использования ленивых вычислений	57
Заключение	59
 Глава 2. Установка и начало работы	 61
Настройка рабочего окружения	61
Загрузка файлов проекта	62
Установка утилиты uv	62

Установка проекта	63
Работа в виртуальном окружении	63
Проверка установки	64
Экспресс-курс по JupyterLab	65
Горячие клавиши в Jupyter	65
Установка Polars в другие проекты	67
Дополнительные зависимости	68
Зависимости для совместимости	68
Зависимости для работы с электронными таблицами	68
Зависимости для работы с базами данных	69
Зависимости для работы с удаленными файловыми системами	69
Зависимости для работы с другими форматами ввода/вывода	69
Зависимости для использования расширенного функционала	70
Установка дополнительных зависимостей	70
Конфигурирование Polars	71
Временная конфигурация с использованием контекстного менеджера	71
Локальная конфигурация с использованием декораторов	74
Компилирование Polars с нуля	75
Особый случай: работа с очень большими наборами данных	76
Особый случай: процессоры без поддержки AVX	76
Заключение	76

Глава 3. От pandas к Polars

Набор данных с животными	78
Сходства pandas и Polars	78
Внешний вид pandas и Polars	79
Разница в коде	79
Разница в выводе	80
Концепции, от которых придется отучиться	85
Индексы	85
Оси	87
Индексирование и срезы	88
Жадность	90
Вседозволенность	93
Синтаксис, который придется забыть	94
Сравнение некоторых операций	94
Из Polars в pandas и обратно	99
Заключение	100

Глава 4. Структуры и форматы данных

Series, DataFrame и LazyFrame	102
Типы данных	104
Вложенные типы данных	106
Пропущенные значения	108
Преобразование типов данных	113
Заключение	116

Глава 5. Жадный и ленивый API	117
Жадный API: датафреймы	117
Ленивый API: ленивые датафреймы.....	119
Разница в быстродействии.....	120
Разница в функционале	121
Атрибуты	122
Методы агрегации	122
Вычислительные методы.....	123
Описательные методы	123
Методы группировки	124
Методы экспорта	124
Методы манипуляции и отбора данных	125
Прочие методы.....	127
Полезные советы.....	128
Переход от ленивых датафреймов к обычным и обратно.....	128
Объединение обычных и ленивых датафреймов	129
Кеширование промежуточных результатов	130
Заключение	131
 Глава 6. Чтение и запись данных	132
Обзор форматов файлов	133
Чтение файлов CSV	133
Корректная обработка пропущенных значений.....	135
Чтение файлов с кодировкой, отличной от UTF-8	137
Чтение данных из Excel.....	139
Работа с несколькими файлами.....	141
Чтение из файлов Parquet.....	143
Чтение JSON и NDJSON.....	145
JSON	145
NDJSON	147
Другие форматы файлов	150
Запросы к базам данных.....	151
Запись данных	154
Запись в формате CSV	154
Запись в формате Excel	154
Запись в формате Parquet.....	155
Другие варианты хранения	156
Заключение	156
 Глава 7. Введение в выражения	158
Методы и пространства имен	160
Выражения в примерах.....	160
Выбор столбцов при помощи выражений	161
Создание новых столбцов с помощью выражений	162
Фильтрация строк при помощи выражений	163

Агрегирование данных при помощи выражений.....	164
Сортировка строк при помощи выражений.....	165
Определение выражения.....	166
Свойства выражений.....	169
Создание выражений.....	171
На основе существующих столбцов.....	172
На основе литеральных значений.....	173
На основе диапазонов.....	176
Другие функции для создания измерений.....	177
Переименование выражений.....	178
Выражения – характерная черта Polars.....	180
Заключение.....	182

Глава 8. Продолжаем знакомиться с выражениями..... 183

Типы операций.....	184
Пример А: поэлементные операции.....	184
Пример В: операции агрегации в одну строку.....	185
Пример С: операции агрегации в одну или несколько строк.....	186
Пример D: операции, расширяющие исходный объект.....	186
Поэлементные операции.....	187
Операции для выполнения математических преобразований.....	188
Тригонометрические операции.....	189
Операции для округления и разбиения на категории.....	190
Операции для работы с пропущенными или бесконечными значениями.....	192
Прочие операции.....	194
Не снижающие размерность операции, применяющиеся к объектам Series.....	195
Накопительные операции.....	195
Операции заполнения и смещения.....	196
Операции, связанные с дублирующимися значениями.....	198
Операции для расчета скользящих показателей.....	199
Операции сортировки.....	200
Прочие операции.....	202
Применяющиеся к объектам Series операции, снижающие размерность до одной строки.....	203
Операции, использующие кванторы.....	204
Операции, вычисляющие статистику.....	205
Операции подсчета.....	206
Прочие операции.....	209
Применяющиеся к объектам Series операции, снижающие размерность до одной или нескольких строк.....	210
Операции, связанные с уникальными значениями.....	210
Операции отбора.....	211
Операции по удалению пропущенных значений.....	212
Прочие операции.....	213

Применяющиеся к объектам Series операции, увеличивающие размерность	216
Заключение	217

Глава 9. Комбинирование выражений..... 218

Встраиваемые операторы против методов.....	219
Арифметические операторы	220
Операторы сравнения	222
Операторы булевой алгебры	226
Битовые операции	228
Использование функций.....	230
When, Then, Otherwise	234
Заключение	236

Глава 10. Выбор и создание столбцов 238

Выбор столбцов	240
Знакомство с селекторами	241
Выбор столбцов по имени	242
Выбор столбцов по типу данных	244
Выбор столбцов по позиции	246
Комбинирование селекторов	248
Создание столбцов.....	250
Операции для работы со столбцами	255
Удаление столбцов.....	255
Переименование столбцов	256
Компоновка столбцов	257
Добавление индексов строк	258
Заключение	258

Глава 11. Фильтрация и сортировка строк..... 259

Фильтрация строк	260
Фильтрация на основе выражений	261
Фильтрация на основе имен столбцов	262
Фильтрация на основе ограничений	263
Сортировка строк.....	264
Сортировка на основе одного столбца	264
Сортировка в обратном порядке	265
Сортировка на основе нескольких столбцов	266
Сортировка на основе выражений	267
Сортировка вложенных типов данных.....	268
Операции по работе со строками	270
Фильтрация пропущенных значений.....	270
Срезы	271
Верхние и нижние	272
Семплирование	272

Полусоединения	273
Заключение	274

Глава 12. Работа с текстовыми, временными и вложенными

типами данных	275
Тип данных String	276
Методы типа данных String	276
Примеры работы со строками	279
Тип данных Categorical	282
Методы для работы с типом Categorical	283
Примеры работы с типом Categorical	283
Тип данных Enum	287
Типы данных, связанные с датой и временем	288
Методы для работы с календарными типами данных	289
Примеры применения типов данных, связанных с датой и временем	291
Тип данных List	294
Методы типа данных List	295
Примеры работы с типом List	296
Тип данных Array	299
Методы типа данных Array	299
Примеры работы с типом Array	300
Тип данных Struct	301
Методы типа данных Struct	302
Примеры работы с типом Struct	303
Заключение	306

Глава 13. Группировка и агрегация данных

Разделяем, применяем и объединяем	308
Контекст GroupBy	308
Примеры применения агрегаций	311
Методы повышенной сложности	316
Построчные агрегации	321
Оконные функции	324
Динамическая группировка	325
Скользящие агрегации	327
Передискретизация данных	331
Заключение	333

Глава 14. Объединение и слияние

Объединение данных	334
Стратегии объединения	335
Объединение по нескольким столбцам	339
Проверка объединения	340
Неточное объединение	342
Стратегии неточного объединения	345

Дополнительная тонкая настройка.....	347
Пример: управление маркетинговыми кампаниями	347
Вертикальное и горизонтальное слияние.....	351
Вертикальное слияние	352
Горизонтальное слияние	353
Диагональное слияние	354
Объединяющее слияние	355
Нестрогие виды слияния	358
Стекинг	359
Добавление	360
Расширение	361
Заключение	362

Глава 15. Изменение формы датафреймов

Широкие датафреймы против длинных.....	363
Разворачивание в широкий формат	366
Разворачивание в длинный формат	371
Транспонирование.....	374
Разворачивание в строки.....	376
Партиционирование датафреймов	380
Заключение	383

Глава 16. Визуализация данных

Поездки на велосипедах в Нью-Йорке	387
Встроенные графические возможности на основе Altair	388
Знакомство с Altair	389
Методы пространств имен plot	390
Графический анализ датафреймов	390
Ограничение на размер	393
Визуализация объектов Series	395
Визуализация «как в Pandas» с помощью hvPlot.....	398
Знакомство с hvPlot	398
Первый график.....	399
Методы пространства имен hvPlot	400
Pandas в качестве движка	401
Ручные преобразования	402
Изменение движка hvPlot	403
Вывод точек данных на географической карте.....	404
Комбинирование графиков.....	405
Добавление интерактивных виджетов.....	406
Готовые к публикации графики при помощи plotnine	407
Знакомство с plotnine.....	408
Диаграммы для исследования данных.....	408
Графики для публикации.....	412
Стилизация датафреймов при помощи Great Tables.....	416
Заключение	420

Глава 17. Расширения Polars	422
Пользовательские функции на Python.....	422
Применение функций к элементам.....	423
Применение функций к объектам Series	425
Применение функций к группам	426
Применение функций к выражениям	429
Применение функций к датафреймам и ленивым датафреймам	430
Регистрация своего пространства имен	431
Плагины Polars в Rust	433
Подготовка	433
Анатомия проекта с плагином	433
Плагин	434
Компиляция плагина	436
Оценка быстродействия	436
Регистрация аргументов.....	437
Использование крейта Rust.....	440
Пример: geo	440
Заключение	450
 Глава 18. Внутреннее устройство Polars	451
Архитектура Polars.....	451
Arrow	452
Многопоточные вычисления и операции SIMD.....	455
Хранение строк в памяти.....	456
ChunkedArrays и Series	457
Оптимизация запросов.....	458
Оптимизации уровня сканирования ленивого датафрейма	459
Другие виды оптимизации	461
Проверка выражений	463
Обзор пространства имен Expr.meta	464
Примеры использования пространства имен Expr.meta.....	464
Профилирование Polars	467
Тестирование в Polars.....	469
Сравнение датафреймов и объектов Series	469
Распространенные антипаттерны.....	472
Использование квадратных скобок для выбора столбцов	472
Неправильное использование метода collect()	473
Использование кода на Python в запросах Polars	474
Заклучение	474
 Приложение. Ускорение Polars с помощью графического процессора	476
NVIDIA RAPIDS	477
Установка движка GPU	478
Шаг 1: установка WSL2 на Windows.....	478

Шаг 2: установка Ubuntu Linux на WSL2	479
Шаг 3: установка необходимых пакетов в Ubuntu Linux.....	480
Шаг 4: установка набора инструментов CUDA	480
Шаг 5: установка зависимостей Python.....	481
Шаг 6: проверка установки.....	481
Использование движка GPU в Polars	482
Настройка движка GPU	482
Неподдерживаемые возможности	483
Эталонное тестирование движка GPU в Polars	483
Решения.....	483
Запросы и данные.....	484
Оборудование.....	485
Результаты и обсуждение	486
Будущие планы по использованию графического процессора в Polars	492
Заключение	492
Предметный указатель.....	493

От издательства

Отзывы и пожелания

Мы всегда рады отзывам наших читателей. Расскажите нам, что вы думаете об этой книге – что понравилось или, может быть, не понравилось. Отзывы важны для нас, чтобы выпускать книги, которые будут для вас максимально полезны.

Вы можете написать отзыв на нашем сайте www.dmkpress.com, зайдя на страницу книги и оставив комментарий в разделе «Отзывы и рецензии». Также можно послать письмо главному редактору по адресу dmkpress@gmail.com; при этом укажите название книги в теме письма.

Если вы являетесь экспертом в какой-либо области и заинтересованы в написании новой книги, заполните форму на нашем сайте по адресу http://dmkpress.com/authors/publish_book/ или напишите в издательство по адресу dmkpress@gmail.com.

Список опечаток

Хотя мы приняли все возможные меры для того, чтобы обеспечить высокое качество наших текстов, ошибки все равно случаются. Если вы найдете ошибку в одной из наших книг, мы будем очень благодарны, если вы сообщите о ней главному редактору по адресу dmkpress@gmail.com. Сделав это, вы избавите других читателей от недопонимания и поможете нам улучшить последующие издания этой книги.

Нарушение авторских прав

Пиратство в интернете по-прежнему остается насущной проблемой. Издательство «ДМК Пресс» очень серьезно относится к вопросам защиты авторских прав и лицензирования. Если вы столкнетесь в интернете с незаконной публикацией какой-либо из наших книг, пожалуйста, пришлите нам ссылку на интернет-ресурс, чтобы мы могли применить санкции.

Ссылку на подозрительные материалы можно прислать по адресу электронной почты dmkpress@gmail.com.

Мы высоко ценим любую помощь по защите наших авторов, благодаря которой мы можем предоставлять вам качественные материалы.

Отзывы о книге

В этой книге Йерун и Тейс проделали блестящую работу, не только рассказав читателям обо всех нюансах Polars, но также поведав о том, как легче расстаться с привычками, приобретенными при использовании других библиотек, таких как `pandas`. Им действительно удалось раскрыть весь потенциал выражений, лежащих в основе эффективного использования Polars и более декларативного и функционального подхода к обработке данных. Уверен, читая эту книгу, вы откроете для себя все глубины Polars и сможете взглянуть на процесс анализа данных по-новому.

— Ричи Винк (*Ritchie Vink*), создатель Polars
(выдержка из вступительного слова)

Библиотека Polars предстала в виде восходящей звезды на горизонте экосистемы обработки и анализа данных в Python и открыла новые границы возможностей для датафреймов следующего поколения. Йерун и Тейс очень вовремя представили общественности свою книгу, позволяющую постичь все нюансы и преимущества Polars.

— Уэс Маккини (*Wes McKinney*),
создатель `pandas` и главный архитектор `Posit PBC`

Библиотека Polars принесла великое множество долгожданных инноваций в мир обработки данных при помощи датафреймов в виде хорошо организованного интерфейса и эффективной реализации. Это позволило вывести процесс обработки и анализа данных в Python на новые, ранее недостижимые высоты. Мы также очень рады тому, что Ричи и его команда являются частью амстердамской экосистемы анализа данных. Меня крайне впечатляют способности Йеруна простым и понятным языком преподносить сложный материал из области науки о данных, будь то работа в командной строке или где бы то ни было еще. Их с Тейсом книга стала ярким подтверждением этого тезиса, и я горячо рекомендую ее к прочтению всем в сообществе анализа данных.

— Ханнес Мюхлеусен (*Hannes Mühleisen*),
партнер по созданию `DuckDB`

Работая совместно с Йеруном и Тейсом над миграцией конвейеров данных в Polars, мы изначально относились к этому проекту с долей скепсиса, но очень быстро убедились в быстродействии и интуитивности этой библиотеки и ее API. И пока авторы допоздна трудились над окончанием своей книги, мы делали все, чтобы улучшить наши рабочие процессы. Мы надеемся, что эта

книга поможет вам в освоении всех тонкостей работы с библиотекой Polars – такой прекрасной и такой ленивой!

— *Марникс ван Лисхаут (Marnix van Lieshout)*
и *Брам Тиммерс (Bram Timmers)*,
специалисты по работе с данными в компании *Alliander*

Эта книга способна изменить ваш подход к анализу данных. Йерун и Тейс проделали феноменальную работу, тщательно выполнив все сравнения и включив необходимые диаграммы и примеры. Polars обладает очень богатым функционалом, и авторы очень бережно все разложили по полочкам. Мне импонирует их стремление визуализировать все анализируемые в книге данные с помощью привлекательных таблиц и графиков!

— *Майкл Чоу (Michael Chow)*,
главный разработчик в компании *Posit PBC*,
специалист по внедрению и поддержке *Great Tables*

Эта книга очень элегантно приоткрывает покров таинственности, коим славится мощная экосистема Polars. В ней Тейс и Йерун в предельно понятной манере объясняют все теоретические основы библиотеки и сопровождают объяснения полезными примерами, что позволяет погрузиться в сложные концепции, не делая материал неподъемным. Работали ли вы ранее с *pandas* или только начинаете свой путь в анализе данных в *Python*, эта книга даст вам все необходимое для создания и осознания рабочих процессов, которые могут выполняться на предельно высокой скорости.

— *Хелла Хаанстра (Hella Haanstra)*,
инженер машинного обучения в компании *Xomnia*

С библиотекой Polars я познакомился довольно давно и даже написал когда-то статью о применении метода *pipe()* к датафреймам. Меня впечатлила скорость работы этой библиотеки, но еще больше я был поражен ее API. Мне показался такой изящный подход очень прогрессивным. С тех пор прошло несколько лет, и Polars превратилась в полноценный инструмент анализа данных. Очевидно, что начинающим разработчикам понадобилось всеобъемлющее пособие по этой библиотеке. И именно его вы сейчас и держите в руках! И пособие, и справочник!

— *Винсент Д. Вармердам (Vincent D. Warmerdam)*,
специалист по работе с данными, сооснователь проекта *Calmcode*

В книге *Python Polars. Подробное руководство* понятным языком рассказывается обо всех особенностях и нюансах работы с Polars. Вместе с тем она содержит огромное множество примеров практического применения этой библиотеки. Поистине отличный источник полезной информации!

— *Стин де Гозейер (Stijn de Gooijer)*,
ведущий разработчик *Polars*

Polars успел стать одним из ведущих фреймворков для работы с данными в Python, особенно в сфере анализа временных рядов и прогнозирования. В настоящее время он полностью интегрирован с такими библиотеками, как MLForecast и StatsForecast от Nixtla, позволяющими эффективно и гибко прогнозировать данные.

Йерун и Тейс проделали потрясающую работу, написав книгу, которая в равной степени подойдет как новичкам в мире анализа данных, так и профессионалам, использующим *pandas*, но желающим развиваться и двигаться дальше.

— Рами Криспин (*Rami Krispin*),
старший менеджер отдела науки о данных и инженерии
в компании Apple

Меня впечатлил уровень погружения в материал для написания такой феноменальной книги, проявленный Йеруном и Тейсом. Я видел несколько книг по Polars, но эта, без сомнений, лучшая! В ней авторы не просто повторяют текст документации, а подробно описывают используемые выражения и в дружеской манере сравнивают эту библиотеку с другими инструментами. Хороший пример такого подхода – описание того, как библиотека Polars из коробки работает с плагином геокодирования. Знакомы вы с Polars или нет, я в любом случае рекомендую вам прочитать эту книгу!

— Марко Горелли (*Marco Gorelli*),
старший инженер-программист в компании Quansight,
ведущий разработчик Polars и *pandas*, создатель пакета *narwhals*

О переводчике



Александр Гинько, обладающий богатым опытом работы в сфере ИТ и более 15 лет посвятивший переводам книг и статей на самые разные темы, в последние годы специализируется на переводе книг в области бизнес-аналитики и программирования для издательства «ДМК Пресс» по направлениям Python, SQL, машинное и глубокое обучение, статистика, Power BI, DAX, Excel, Power Query, Tableau, R... На данный момент в активе Александра уже более

25 книг, включая одну авторскую, и он продолжает плодотворно работать над переводом и написанием новых книг.

Если вы читаете эту книгу, возможно, вам также будут интересны книги «Сверхбыстрый Python» (<https://dmkpress.com/catalog/computer/programming/python/978-5-93700-226-6>), «Python: практическое руководство по Pandas (200 упражнений)» (<https://dmkpress.com/catalog/computer/programming/python/978-5-93700-227-3>), «Введение в статистическое обучение с примерами на Python» (<https://dmkpress.com/catalog/computer/statistics/978-5-93700-217-4>) и «Инженерия данных в Python: основы анализа данных с помощью Pandas, NumPy и Scikit-Learn» (<https://dmkpress.com/catalog/computer/programming/python/978-5-93700-381-2>) в переводе Александра.

Помимо перевода книг, Александр ведет свой канал в Telegram (https://t.me/alexanderginko_books), на котором вы можете из первых уст получить ответы на все интересующие вас вопросы об уже переведенных книгах, находящихся в работе и запланированных на будущее. Также на канале можно найти промокоды на все книги Александра для покупки книг на сайте издательства «ДМК Пресс» с большими скидками. Купить книги Александра и следить за переводом новых книг в режиме реального времени можно с помощью его бота в Telegram по адресу https://t.me/alexanderginko_books_bot.

Вступительное слово от создателя Polars Ричи Винка

Это никогда не должно было оказаться правдой, но...

В декабре 2019 года я стал отцом, после чего практически сразу началась пандемия COVID-19. Признаюсь, эти события меня прилично выбили из колеи. Я пытался не потерять рассудок и хоть как-то совмещать непривычный для меня ритм жизни новорожденного ребенка с тщетными попытками удержаться за свое собственное я. В очередном приступе родительской бессонницы я решил, что мне срочно нужен какой-то новый проект. Если честно, это было только ради развлечения.

На работе, программируя на Rust, я столкнулся с задачей объединения данных из двух файлов CSV, и оказалось, что сделать это не так-то просто. Тогда я задумался, а есть ли способы облегчения этой задачи без использования SQLite. Как и принято у программистов, очень скоро я сел за написание собственного алгоритма объединения.

В то время у меня было не так много опыта программирования на Rust, и я мало что знал об оптимизации. В общем, несмотря на распирающую меня гордость от написания своего первого алгоритма объединения данных, я не мог не признать, что по эффективности он значительно уступал аналогичной операции, выполненной в pandas. Пришедшее на смену гордости разочарование и стало тем заветным семенем, из которого по прошествии времени возник Polars.

Та моя неудача поспособствовала тому, что я начал глубоко изучать движки баз данных и погружаться в методы оптимизации, заложенные в языке Rust, что вылилось в полтора года бесконечных проб и ошибок. В процессе понимания того, как работают базы данных, какие в них используются алгоритмы, как устроена работа с памятью и т. д. мои изначальные цели изменились. Мне стало недостаточно просто опередить pandas при объединении данных, я захотел написать полноценный пакет для работы с датафреймами для Rust. Это вылилось в желание создать свой собственный высокоэффективный движок запросов, который мог бы успешно соперничать с лучшими представителями семейства движков в Python.

Свои устремления я основывал на преимуществах и недостатках pandas, декларативных подходах, использующихся в SQL и PySpark, принципах функ-

ционального программирования и строгости системы типов, характерной для Rust. Поначалу я думал, что мой движок будет основан на API pandas. Но очень быстро понял, что подобные ограничения могут негативно сказаться как на моем творческом посыле сделать что-то новое, так и на эффективности итогового продукта. Избавившись от этих сдерживающих меня пут и объединив жадный и ленивый интерфейсы в один API на основе выражений, я получил зародыш того, что сегодня именуется библиотекой Polars.

15 марта 2021 года я выложил свое творение на PyPI в качестве исследовательского проекта. После этого я постепенно превратил свое детище в полноценный пакет для работы с датафреймами. Успех проекта позволил мне в результате заручиться достаточной финансовой поддержкой и основать собственную компанию под названием Polars Inc.

Сегодня Polars – это самостоятельная компания, которой я безмерно горжусь и которая является моим нескончаемым источником энергии. Наконец-то все мои идеи могут быть воплощены в жизнь (ну, ладно, не все, но я преисполнен амбиций и не теряю надежд). В глобальном смысле в Polars я хочу воплотить идею создания единого API, который будет работать повсеместно на одном и том же стеке технологий вне зависимости от ваших требований к объему обрабатываемых данных.

Вместе с моей преданной командой в Polars Inc и постоянно растущим сообществом энтузиастов мы можем решить великое множество задумок. Каждый проект предполагает свои собственные сложности, и поиск универсальных решений для сценариев разных масштабов – дело невероятно трудное, но интересное. Одна из таких сложностей связана с обработкой временных зон. Это может звучать как пустяковая задача, но на самом деле преобразование часовых поясов с сохранением консистентности данных из разных систем – дело не из легких. Еще одна техническая сложность, с которой мы в настоящее время столкнулись, связана с преобразованием API в полноценную потоковую модель. Этот процесс требует полного переосмысления того, как различные операции, например вычисление среднего значения или доступ к последней строке, должны выполняться в контексте потока в сравнении с пакетной обработкой с хранением данных в оперативной памяти. Я уверен, что мы справимся с этими трудностями.

С момента появления библиотеки Polars ее рост превысил любые мои ожидания. Для меня было большой неожиданностью увидеть, в каких сферах применяется наше творение. К примеру, это касается его использования при симуляции метода конечных элементов с участием полутора тысяч объединений. Также я видел, что Polars используют в задачах, связанных с метапрограммированием, например при генерировании сложных запросов, которые проблематично написать вручную.

В будущее Polars я смотрю с большим воодушевлением. В первую очередь мы нацеливаемся на расширение потоковых возможностей, что позволит вам комфортно работать с огромными наборами данных на своих ноутбуках. Мы также работаем над созданием распределенной облачной среды, которая поможет быстрее выполнять большие вычисления. Кроме того, в наши планы

Конец ознакомительного фрагмента.

Приобрести книгу можно

в интернет-магазине

«Электронный универс»

e-Univers.ru